Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies

Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies

James P. Kirby

School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, Scotland (U.K.)

Abstract

Unlike many languages of Southeast Asia, Khmer (Cambodian) is not a tone language. However, in the colloquial speech of the capital Phnom Penh, /r/ is lost in onsets, reportedly supplanted by a range of other acoustic cues such as aspiration, a falling- or low-rising f0 contour, breathy voice quality, and in some cases diphthongization, e.g. /krɑː/ 'poor' > [kɔ̀ɑ], [kʰɔ̌ɑ], [kɔ̤ɑ̤], /kruː/ 'teacher' > [kùː], [kʰǔː], [kṵː]. This paper presents the results of production and perception studies designed to shed light on this unusual sound change. Acoustic evidence shows that colloquial /CrV/ forms differ from reading pronunciation forms in terms of VOT, f0, and spectral balance measures, while a pair of perceptual studies demonstrate that f0 is a sufficient cue for listeners to distinguish underlying /CrV/-initial from /CV/-initial forms, but that F1 is not. I suggest that this sound change may have arisen via the perceptual reanalysis of changes in spectral balance, coupled with the coarticulatory influence of the dorsal gesture for /r/.

Key words: Khmer; tonogenesis; voice quality; sound change; phonologization

*Tel: +44 (0)131 650 3952; fax: +44 (0)131 651 3190.
Email address: j.kirby@ed.ac.uk (James P. Kirby)

1. Introduction

Khmer, the national language of Cambodia, is notable for being a non-tonal language in what may be the most 'tone-prone' area of the world (Matisoff, 1973). However, an incipient tone contrast has been reported in the colloquial speech of the capital Phnom Penh (henceforth PP) for at least 50 years, albeit in a highly restricted context: in words containing /r/ in onset position such as /kruː/ 'teacher' or /rien/ 'to learn', the /r/ is reportedly supplanted by aspiration and some type of f0 contour, resulting in pronunciations such as [kʰŭː] or [hĭen]. Since Khmer contrasts plain and aspirated stops in onsets[1] (see Table 1), this suggests that the contrast in PP between forms like /kruː/ 'teacher' and /kʰuː/ 'old', or between /rien/ 'to learn' and /hien/ 'to dare', may now be maintained colloquially by a difference in pitch. Setting aside the analytic issue of whether or not this constitutes a phonemic use of f0, it is extremely interesting for researchers interested in the process of tonogenesis, as such a pathway has not been described for any other language.

While this phenomenon has been noted in the literature for some time, existing descriptions are inconsistent. The earliest accounts (Noss, 1966; Huffman, 1967) describe a (low) rising tone and a deleted or spirantized trill, e.g. /kruː/ > PP [kúː], /rien/ > PP [híen]. Pisitpanporn (1994, 1999) reports loss of /r/ in PP as accompanied by increased post-release aspiration in addition to a falling-

[1] Despite the existence of a large number of minimal pairs supporting a classical phonemic distinction between aspirated and unaspirated voiceless initial stops, traditional accounts (e.g. Henderson, 1952; Huffman, 1967; Pinnow, 1980) analyse these forms as a sequence of stop + /h/, citing alternations such as /tʰom/ 'big' ∼ /tomhom/ 'size' or /kʰəŋ/ 'angry' ∼ /kɑmhəŋ/ 'anger' as evidence for the separability of the stop and /h/ by the nominalising infix *-Vm-*.

Table 1: Three types of onsets in Standard Khmer: plain stops, aspirated stops, and stop + /r/ clusters.

/piː/	'two'	/pʰiː/	'grandly'	/priː/	'plug'
/taː/	'grandfather'	/tʰaː/	'to say'	/traː/	'seal, stamp'
/kuː/	'pair'	/kʰuː/	'old'	/kruː/	'teacher'

rising pitch contour, transcribing the PP pronunciations of Standard Khmer forms like /kruː/ as [kʰŭː]. Post-release aspiration of colloquial /CrV/ forms was also found by Wayland & Guion (2005) in an acoustic study of the speech of two male speakers of PP Khmer. While one speaker displayed a falling-rising f0 contour, the other showed low-rising f0 in colloquial forms.

In addition to pitch and aspiration, vowel quality differences may also accompany loss of /r/ in PP Khmer. Forms containing the low vowels /a aː ɑː/ have been described as being produced with diphthongs [ea] and [ɔɑ], e.g. /traː/ 'seal, stamp' $>$ PP [tʰĕa], /krɑː/ 'poor' $>$ PP [kʰɔ̆ɑ], while the diphthong /ae/ may be produced as a monophthong /eː/ or /ɛː/, e.g. /sraek/ 'shout' $>$ PP [sĕːk] (Pisitpanporn, 1994, 1999; Wayland & Guion, 2005). Filippi & Vicheth (2009) transcribe PP variants of Standard Khmer forms such as /krɑː/ and /kruː/ as [k̤o̤a̤] and [k̤ṳː], suggesting that breathy voice quality, previously an important feature of Khmer pronunciation and one still preserved in some conservative dialects (Wayland & Jongman, 2003), may also play a role in colloquial PP Khmer.

Thus, the first goal of the present study is to provide a thorough analysis of the acoustic realization of both standard and colloquial PP Khmer /CrV/ forms, in order to compare their acoustic properties to one another as well as to /CV/ and /CʰV/ forms. The focus here is on acoustic measures of aspiration (VOT),

pitch (f0), vowel quality (F1/F2), voice quality ($H1^{*}-H2^{*}$ and $H1^{*}-A3^{*}$), and the realization of /r/.

The second goal is to assess the perceptual salience of likely cues to colloquial /CrV/ forms. While words such as /kɑː/ 'neck' and /krɑː/ 'poor' may indeed be produced with differences in fundamental frequency and/or vowel height, it has not been demonstrated that the magnitudes of these differences are actually salient to native listeners. In order to determine whether, and to what degree, listeners can use differences in f0 or F1/F2 to identify colloquial /CrV/ forms, a pair of perceptual identification task was conducted in which f0 or F1/F2 was systematically manipulated along with other potentially salient acoustic cues.

The third goal of this study is to use the production and perception findings to assess potential explanations for this unusual sound change. Two possibilities are considered here. Wayland & Guion (2005) suggest an articulatory-aerodynamic explanation, based on the cross-linguistic tendency for trills to devoice in faster or colloquial speech (McGowan, 1992; Solé, 2002). Their idea is that the high rate of airflow necessary for trilling may have led to a higher rate of vocal fold vibration at the onset of the vowel, subsequently conditioning a greater drop in f0 relative to forms not containing a trill. If the trill were to devoice, this increased airflow may have then been reinterpreted either as /h/ or as aspiration on the preceding consonant, and the resulting pitch contour could have been reinterpreted as a tonal feature of the vowel. The diphthongization of low vowels is hypothesized to be a coarticulatory effect of the trill, presumably motivated by the observation that the dorsal gesture associated with /r/ results in a raised tongue body resembling vowels such as /ʌ/ or /ɔ/ (Gick et al., 2002, 2006).

Variation in the realization of /r/ raises another possibility, however. Trills

are often found to vary cross-linguistically with taps, approximants, and especially fricatives, both voiced and voiceless (Ladefoged & Maddieson, 1996; Solé, 2002). If the resulting frication noise were to extend into the onset of the nucleus, this could cause vowels in /CrV/ contexts to be perceived as breathy (Klatt & Klatt, 1990). Breathiness frequently co-occurs with low f0 (Ohala, 1973) and can perceptually condition F1 lowering (Lotto et al., 1997); as such, it has often been implicated in the diachronic development of both diphthongization patterns as well as 'register' systems, where a language's vowels are separable into groups differing in height, pitch, and/or voice quality (Huffman, 1976; Denning, 1989). The process by which Khmer acquired its sizeable inventory of vowels almost certainly involved a stage of contrastive voice quality (Henderson, 1952; Ferlus, 1992; Wayland & Jongman, 2002), and some scholars have proposed that the canonical path to tone is similarly mediated by phonation type contrasts (Pulleyblank, 1978; Diffloth, 1989; Thurgood, 2002). If fortition or devoicing of /r/ does condition the percept of breathy phonation, the development of both f0-based contrast and diphthongization would be consistent with this more general hypothesis.

One means of assessing these hypotheses is to look for evidence of the relevant phonetic precursors in standard/careful productions of /CrV/ forms. The articulatory-aerodynamic hypothesis predicts that, if the source of the pitch contour and diphthongization in colloquial forms are coarticulatory effects induced by /r/, these effects should be present to a greater degree in standard /CrV/ forms compared to /CV/ and /C^{h}V/ forms. The breathiness hypothesis is more difficult to test directly. However, synchronic variation in the realization of /r/, or evidence that voice quality plays a significant role in the production and/or perception of colloquial /CrV/ forms, would be consistent with such an evolutionary trajectory.

The remainder of the paper is organised as follows. Section 2 provides the relevant details of Khmer phonetics and phonology. Section 3 presents the results of the production study. Colloquial PP forms are found to be characterized by increased post-release aspiration and breathy phonation in addition to previously described differences in f0 and F1. Furthermore, considerable variation in the realization of /r/ is also observed, including frication as well as devoicing. Section 4 demonstrates that f0 is a sufficient cue for listeners to distinguish underlying /CrV/ from /CV/ forms, but that other cues may serve to enhance this contrast. Conversely, listeners do not appear to use variation in F1 as a cue to lexical identity. Section 5 discusses the implications of these findings in the broader context of the evolution of tone and register systems.

2. Khmer language

Khmer is the sole member of the Khmeric branch of the Mon-Khmer group of Austroasiatic languages. It is spoken natively by around 13 million people in Cambodia as well as over a million ethnic Khmers in Vietnam. Northern (Surin) Khmer, spoken in Thailand, claims at least another half million speakers (Diffloth, 2003). There are also sizeable Khmer diaspora in Canada, China, France, Laos, and the United States.

Most varieties of Khmer have the consonant inventory /p b m w t d s n l r c ɲ j k ŋ h ʔ/ (and, arguably, an aspirated stop series /p^h t^h c^h k^h/: see Diffloth, 2003 and footnote 1 above). Of these, only /b d r s/ are prohibited word-finally. Implosion of /b d/ $>$ [ɓ ɗ], while possibly canonical, is non-contrastive. While we focus here only on /Cr-/ clusters, Khmer admits a large number of complex onsets (Huffman, 1972).

/r/ has alternately been described as a 'lingual roll' (Henderson, 1952), a retroflex flap (Huffman, 1970), and an alveolar trill (Huffman, 1967; Wayland & Guion, 2005). The palatal plosive /c/ is generally realized as a palato-alveolar affricate [t͡ɕ] (Wayland & Guion, 2005). The realisation of the semivowel /w/ may vary considerably: Henderson (1952) reports variation between [w], [v] and [ʋ], while Filippi & Vicheth (2009) transcribe it as [β̞] in onsets but as [i̯] in codas.

The vowel inventory of Khmer is a matter of considerable debate and varies with dialect, but proposals are invariably large (Huffman, 1970; Headley et al., 1997; Diffloth, 2003). Here we adopt the notation of Huffman (1970), who suggests a minimally accurate system must contain no fewer than 31 nuclei: 10 long vowels /iː eː ɛː ɨː əː aː ɑː uː oː ɔː/, 8 short vowels /i e ɨ ə a ɑ u o/, 10 long diphthongs /iə ɨə uə ei əɨ ou ae aə ao ɔə/ and 3 short diphthongs /ĕə ŭə ŏə/.

As noted above, Khmer vowels are usually described as falling into one of two 'registers', the consequence of a historical process of initial obstruent devoicing which resulted in a doubling of the previous vowel inventory (Pinnow, 1980). The original laryngeal contrasts are thought to have been transferred to the following vowels via a stage of contrastive voice quality, the effects of which can still be observed in certain conservative Khmer dialects (Wayland & Jongman, 2003).

3. Production study

Wayland & Guion (2005) examined three acoustic parameters distinguishing colloquial from reading /CrV/ forms: degree of medial f0 drop, voice onset time (or fricative noise duration in the case of the voiceless affricated stop /c/ $\sim$ [t͡ɕ]), and vowel quality (F1 and F2). They found significant differences by condition for all of these parameters for the two speakers they recorded. However, their

wordlist consisted solely of /CrV/ items such as /krɑː/ ‘poor’, so it is not clear how, or whether, colloquial /CrV/ productions differ from /CV/ or /C^{h}V/ productions. Thus, the present study employed an expanded reading condition wordlist in order to facilitate such a comparison.

The findings speak to three hypotheses raised by Wayland & Guion (2005). The first hypothesis regards the source of the falling-rising f0 contour. Wayland & Guion suggest that the high volume of translingual airflow necessary to produce a trill could condition a higher f0 during the trill itself, creating an overall greater drop in f0 at the onset of /CrV/ forms compared to /CV/ forms (2005:62). They found preliminary support for this hypothesis in an acoustic and aerodynamic study of Thai (Guion & Wayland, 2004). If this is the case in PP Khmer, we might expect to observe such a drop in reading pronunciations of /CrV/ forms, inasmuch as they can be assumed to represent an earlier stage of the colloquial pronunciations.

The authors further hypothesized that diphthongization (F1 lowering) could result due to the lingual coarticulatory effects of the trill on the following (low) vowel. Again, if this is the case, we should find lower F1 at vowel onset in reading /CrV/ compared to /CV/ forms, especially for low vowels, where the difference in size and location of the oral constriction will be greatest.

Finally, Wayland & Guion suggest that the aspiration observed in colloquial /CrV/ forms could be traced to a devoiced trill, which may have been perceptually reanalyzed as aspiration on the preceding consonant. This leads to the expectation that trills in reading pronunciations of /CrV/ forms, or in colloquial /CrV/ forms where the trill is not dropped, may be produced as devoiced, fricated and/or otherwise fortited. This hypothesis was tested by examining the acoustic realization

of the trill in careful (read) speech.

3.1. Materials and methods

A wordlist of 41 lexical items was constructed (see Appendix A). The list was constrained by the desire to include minimal triplets, as well as by gaps in the Khmer lexicon; thus, it was not possible to perfectly balance items for segmental content or syllable shape. A Khmer assistant confirmed the lexical status of all items along with their meanings and spellings (although several of the items admit alternate spellings).

3.1.1. Participants

20 native speakers of PP Khmer (13 female), aged 21 to 60, participated in the present study. Participants were mainly born in Phnom Penh, although some had moved to the city at a young age. Their level of education was variable, but all were literate in Khmer and many had some post-secondary (university-level) education. 5 participants spoke some English and 1 French; the remaining 14 participants reported themselves as monolingual in Khmer. A Khmer assistant explained the procedure to the participants and answered questions. Participants received a small stipend for their participation in both the production and perception tasks.

3.1.2. Procedure

The production segment consisted of a self-paced *reading* task followed by a *repetition* task. In the self-paced reading task, participants were presented visually with a target item in the frame sentence [kʰɲom tʰaː ___ tiət] 'I say ___ again'. Participants were instructed to read each sentence aloud at a normal speaking rate. Each of the 41 target items appeared 3 times, in randomized order.

Colloquial variants were elicited using a repetition task. While Wayland & Guion (2005) used orthographic prompts to elicit both reading and colloquial pronunciations, participants in a pilot experiment were unwilling or unable to produce colloquial variants when presented with visual primes. Instead, a system was devised where the experimenter prompted the participant orally with the Standard Khmer form, whereupon the participant would respond with the colloquial variant[2]. Each of the 15 target items was elicited three times in randomized order.

24 bit, 44.1 kHz audio recordings were made with a Marantz PMD-661 portable solid state recorder and a Beyerdynamic Opus 55.18 Mk II omnidirectional headset condenser microphone fitted with a CV 18 preamplifier. All recordings were made in reasonably quiet rooms at the Buddhist Institute or the author's accommodations in Phnom Penh.

3.2. Temporal and acoustic analysis

While the design potentially provided a total of 168 items per participant (41 reading condition items $\times$ 3 repetitions plus 15 colloquial condition items $\times$ 3 repetitions), vagaries of data collection (e.g. participants accidentally clicking through a trial during the self-paced reading task) led to a slightly smaller number of items recorded for some participants. 4 participants consistently produced forms containing /r/ in both conditions, suggesting the spread of /r/-loss may be variable in the population (although subsequent analysis could not identify predic-

[2]A reviewer raises the possibility that participant responses may have been influenced by non-standard or colloquial cues present in the primes. Post-hoc examination of the primes did not reveal any non-standard pronunciations or cues that would have impacted participant responses. Importantly, since all participants completed the reading task first, they were not exposed to colloquial pronunciations of any of the target items prior to producing them spontaneously.

tors such as age, gender, or education as the source of this effect). Since the focus of this study was on the acoustic differences between /r/-full and /r/-less forms, data from these participants were not analyzed further.

Once recordings were transferred to PC, 2664 tokens were isolated and segmented using Praat 5.2.26 (Boersma & Weenink, 2011); except where noted, the discussion that follows focuses on the 2099 tokens with obstruent onsets, i.e. ignoring /r/- and /sr/-initial pairs. Initial examination revealed that /CrV/ forms in the reading condition often contained an excrescent vocalism intruding between the onset and /r/ (cf. Huffman, 1972); this led to five potential region labels: (o)nset (burst + VOT), excrescent (v)ocalism, t(r)ill/tap (further sub-labeled as voiced or voiceless trill/tap, voiced or voiceless fricative, or approximant), and syllable ri(m)e (sub-segmented into nucleus and (c)oda when appropriate). Onset of the initial obstruent was defined as the duration from the release burst to the first complete vibration of the vocal folds. Onset of /r/ was defined either as coextensive with the end of the onset of the initial obstruent (for forms with no excrescent vowel) or the first period of decreased amplitude vocal fold vibration (for forms where an excrescent vowel was present). In general, all measurements were defined with reference to the waveform, although spectrographic information was also taken into account, especially when segmenting /r/, where the breaking away of the tongue tip gesture tended to leave a clear spectral signature. All interval boundaries were automatically snapped to the nearest zero crossing.

After segmentation, the duration of each labeled interval was recorded, along with spectral measures taken from the nuclear vowel at 11 equally spaced timepoints and, if present, the excrescent vowel (for which a single measurement was taken at segment midpoint). The first and last nuclear vowel measurements were at

least 12.5 ms from the boundary, using a 25 ms analysis window. f0 was measured using the autocorrelation method of Boersma (1993) with a 15 ms frame duration and a 500 Hz pitch ceiling. Formants were measured by computing LPC coefficients using Praat's implementation of the Burg algorithm, using a 25 ms window with pre-emphasis applied from 50 Hz, and then smoothed using the `Track...` function. Harmonic structure was determined through spectral analysis using FFT and long term average spectra applied to 25 ms windows centered at the measurement points. The amplitudes of the first two harmonics (H1 and H2) were measured along with the amplitude of the most prominent harmonic of the third formant (A3) in order to calculate H1−H2 (an indicative measure of open quotient) and H1−A3 (a measure of spectral tilt). These measures were subsequently corrected for the effect of the first two formants on the vocal tract transfer function (see section 3.3.6 below).

3.3. Results

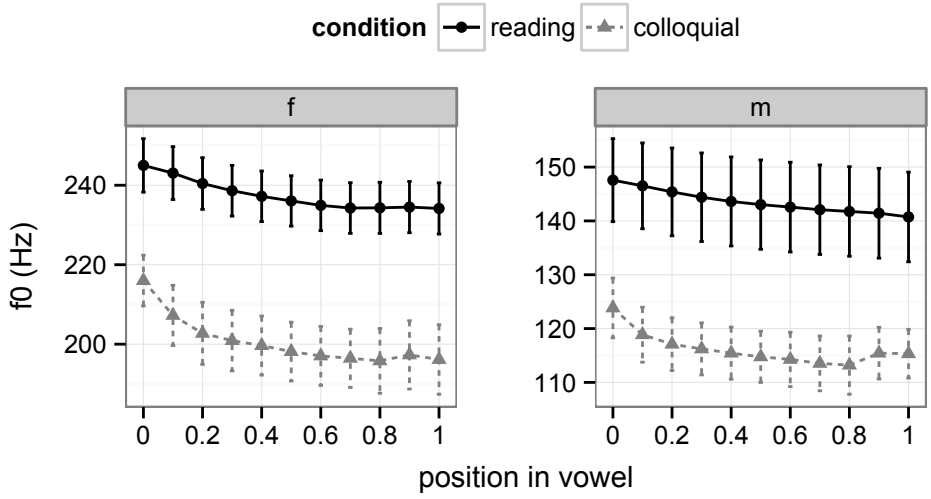
All data were analyzed using generalized linear mixed models (GLMMs) as implemented in the R package `lme4` (Bates et al., 2013; see Appendix B).[3]

3.3.1. Fundamental frequency (f0)

Figure 1 plots f0 (on the Hz scale) over the time course of the vowel by condition for male and female productions of /CrV/ forms. For speakers of both genders, f0 of words produced in the colloquial condition was consistently lower

[3]Note that at present, there is no uncontroversial method for computing p-values for multilevel models incorporating random correlation parameters. Here, I report (potentially anticonservative) p-values obtained by subtracting the number of predictors from the number of observations, and estimating the p-value in the usual fashion (Baayen et al., 2008).

Figure 1: Average f0 (in Hz) by gender (left: female, right: male) and condition for /CrV/ items. Bars show standard error of the mean.



than f0 of words produced in the careful/reading condition at all timepoints.

To estimate the magnitude of the f0 differences, separate GLMMs (using the identity link and variance functions) were fit for male and female talkers. In addition to CONDITION (`reading` or `colloquial`), these models included both linear and nonlinear terms for centered vowel POSITION along with their interaction. Both models also included random intercepts for subjects and items along with correlated by-subject and by-item random slopes for POSITION and CONDITION.[4] Mean f0 in the `colloquial` condition was around 40 Hz lower for female talkers ($\beta = -39.98, SE = 6.22, t = -6.4, p < 0.001$) and 30 Hz lower for male talkers ($\beta = -30.33, SE = 7.47, t = -4.1, p < 0.001$). For female talk-

[4]Additional covariates such as age, gender, education, and (vocalic) register, were considered but did not emerge as significant.

ers, f0 fell throughout the course of the syllable in both conditions (`reading`: $\beta = -2.12, \textit{SE} = 0.32, t = -6.7, p < 0.001$; `colloquial`: $\beta = -1.57, \textit{SE} = 0.26, t = -6.2, p < 0.001$); for males, f0 fell significantly in the `colloquial` condition ($\beta = -1.01, \textit{SE} = 0.21, t = -4.8, p < 0.001$) and marginally in the `reading` condition ($\beta = -0.97, \textit{SE} = 0.52, t = -1.9, p = 0.06$), although the difference between these coefficients is unlikely to be significant.

The shape of the f0 effect observed in these data differs from that reported in Wayland & Guion (2005) (although it is consistent with Thạch Ngọc Minh, 1999, who describes falling f0 on this class of items in a Khmer dialect spoken in neighboring Vietnam). One possibility for this divergence may have to do with the different prosodic contexts (phrase-medial vs. phrase-initial) in which the targets were elicited. Prosodic structure is known to influence the realization of suprasegmental features (Shattuck-Hufnagel & Turk, 1996; Cho, 2011); in particular, global prosodic context has been found to induce final f0 lowering in a variety of languages including English (Liberman & Pierrehumbert, 1984), Japanese (Pierrehumbert & Beckman, 1988), and Mandarin (Shih, 1988). Thus a low-falling contour could potentially result from a falling-rising tone interacting with a L% boundary tone.

The influence of prosodic structure on the realization of lexical f0 has also been found to vary cross-linguistically. For example, while Thai shows evidence for phrase-final boundary tones, they may be overridden by lexical tone targets (Pittayaporn, 2007). Thus, it is unclear whether or not a boundary tone, if present, would have a lowering effect on the lexical tone target, especially given the finding that lexically contrastive properties are preserved under differences in prosodic structure (Nakai et al., 2012). In the absence of a detailed account of Khmer

prosody, we are left to speculate on the nature of the relationship between lexical and post-lexical f0 targets for the time being.

However, examination of the individual production results suggests that the shape of the f0 contours may reflect considerable interspeaker variation. Figure 2 shows f0 in `reading` versus `colloquial` speech for four participants. While a falling-rising pitch contour characterizes the speech of participants s5 and s9, f0 is strictly falling for participant s8 and is realized as low-level for participant s14. These data suggest that the fundamental difference between the realization of /CrV/ forms in careful and casual speech may be one of pitch height, while differences between the present study and previous findings may involve individual variability, either in terms of the lexical pitch targets, the interaction of lexical and post-lexical tones, or both.[5]

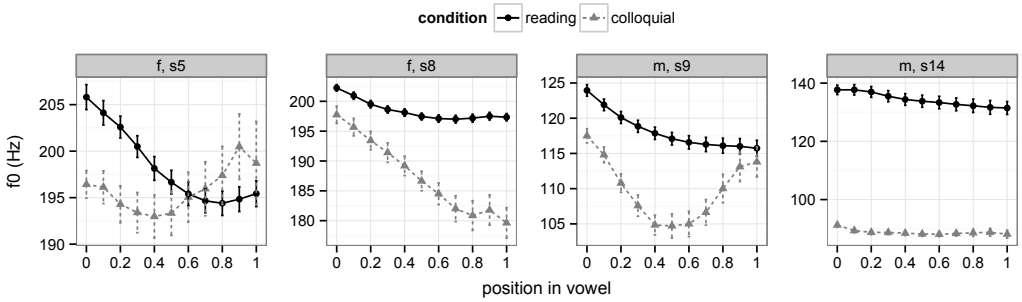
In light of these findings, and for the sake of consistency, colloquial /CrV/ items will hereafter be indicated with a grave accent, e.g. [k^{h}ù:], while recognizing that the precise phonetic variation may vary between speakers and possibly prosodic contexts as well.

3.3.2. f0 drop

To the extent that colloquial PP pronunciations like [k^{h}ù:] can be assumed to derive from Standard Khmer pronunciations like /kru:/, we might expect to find the relevant phonetic precursors to have greater magnitude in forms like /kru:/ compared to /k^{h}u:/ or /ku:/. In particular, if the f0 contour is induced by the aero-

[5]Five participants occasionally produced colloquial (i.e., /r/-less) variants of /CrV/ items in the `reading` condition. An analysis of this small subset of tokens, produced in the same prosodic environment as /r/-ful forms, reveals a pattern of low/falling f0 similar to that in Figure 1.

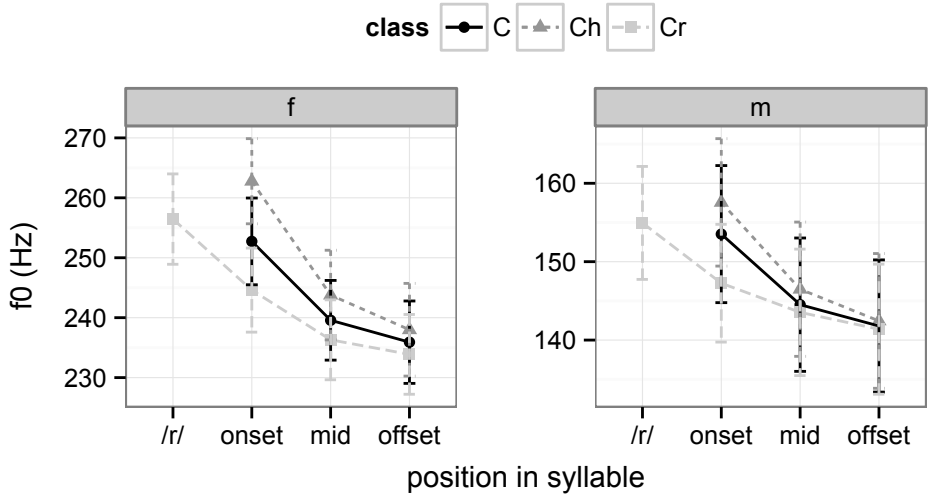
Figure 2: f0 for /CrV/ items, 4 speakers (reading: solid, colloquial: dashed). Bars indicate within-speaker standard error of the mean.



dynamic requirements for trilling, as suggested by Wayland & Guion (2005), the magnitude of this effect may be greater in reading condition /CrV/ forms compared to /CV/ and /C$^{\text{h}}$V/ forms.

This hypothesis was explored by comparing the difference between f0 at voicing onset (i.e., during /r/ or the excrescent vocalism for /CrV/ sequences; at nucleus onset for /CV/ and /C$^{\text{h}}$V/ sequences) and vowel midpoint. To get a better sense of the differences in magnitude of f0 drop across syllable types, a GLMM was fit using POSITION and reverse Helmert coded CLASS (three levels, with /CV/ as the reference level), with random intercepts for subjects and items. The coefficient estimate for CLASS=/C$^{\text{h}}$V/ ($\beta = 16.73, SE = 1.11, t = 15.1, p < 0.001$) indicates that f0 falls slightly more steeply following aspirated stops than unaspirated stops, as might be expected given that f0 tends to be higher following aspirated stops (Hombert, 1975; Lai et al., 2009). In /CrV/ contexts, however, the difference in f0 between onset of voicing to rime midpoint is not significantly different from that of /CV/ syllables ($p = 0.11$).

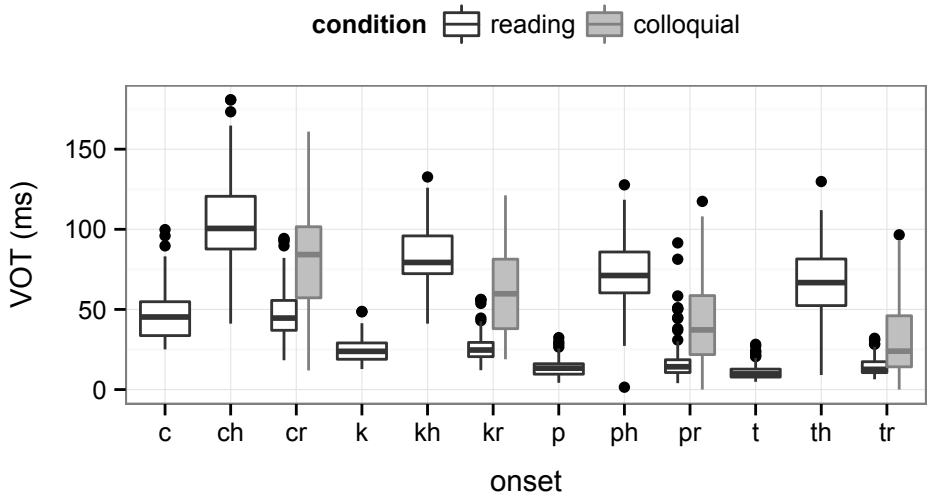
Figure 3: Average f0 (in Hz) by gender (left: female, right: male) and syllable type for `reading` condition items at four timepoints. Bars show standard error of the mean.



3.3.3. Voice onset time

Figure 4 compares the VOT of /CrV/ items in the `colloquial` condition with those of the three classes of items in the `reading` condition. VOT follows the expected distribution by place, with the affricated stop /c/ having the greatest VOT, followed by velar /k/, with /p/ and /t/ having comparable VOT durations. To get a more precise sense of this difference, the interaction of Helmert-coded PLACE OF ARTICULATION (4 levels) and CONDITION (2 levels) was used as a predictor of VOT in a multilevel regression with random intercepts for subjects and items, including by-subject slopes for all three covariates (PLACE, CONDITION, and their interaction). As expected, there was a main effect of CONDITION; VOT for /CrV/ items is on average around 27 ms longer in the `colloquial` condition ($\beta = 27.45, SE = 3.98, t = 6.9, p < 0.001$). This effect is mediated by the inter-

Figure 4: VOT for /CV/, /C^{h}V/ and /CrV/ forms by onset in the `reading` (open) condition contrasted with /CrV/ forms in the `colloquial` (shaded) condition.



action with PLACE: in the `colloquial` condition, VOT for [-anterior] /k/ was slightly longer than [+anterior] /p, t/ ($\beta = 4.46, SE = 1.11, t = 4, p < 0.001$). Conversely, the main effect of PLACE was only significant for the difference between affricated /c/ and the mean of /p t k/ ($\beta = 7.42, SE = 2.7, t = 2.7, p < 0.01$).

Wayland & Guion (2005) reported that onsets in colloquial /CrV/ forms were aspirated (mean VOT $\approx$ 53.5 ms) relative to the corresponding reading pronunciations (mean VOT $\approx$ 12 ms). Although VOT of colloquial /CrV/ forms in the present study were also found to be increased relative to reading pronunciations, mean duration differed from that of reading condition /C^{h}V/ forms. The size of this effect was estimated by fitting a multilevel regression with CLASS (3 levels) and PLACE (4 levels) as covariates, with by-participant random slopes and in-

tercepts for both predictors along with by-item random intercepts. The estimate of average VOT for aspirated stops in `reading` condition ($\beta = 58.06, SE = 3.01, t = 19.3, p < 0.001$) is over three times that of VOT of plain stops in /CrV/ sequences in the `colloquial` condition ($\beta = 16.02, SE = 2.87, t = 5.6, p < 0.001$), although as seen from Figure 4, this effect was also modulated by PLACE.

While this 'intermediate' VOT category seems quite robust across place of articulation and speaker, its source is less clear. One possibility is that the length of VOT in colloquial pronunciations is related to the duration of /r/ in reading pronunciations. To explore this hypothesis, mean (speaker-normalized) duration of aspiration in colloquial /CrV/ forms was used to predict duration of the /r/ in reading forms for the same items in a multilevel regression with by-subject and by-item intercepts. The duration of aspiration predictor was not significant ($p = 0.62$).

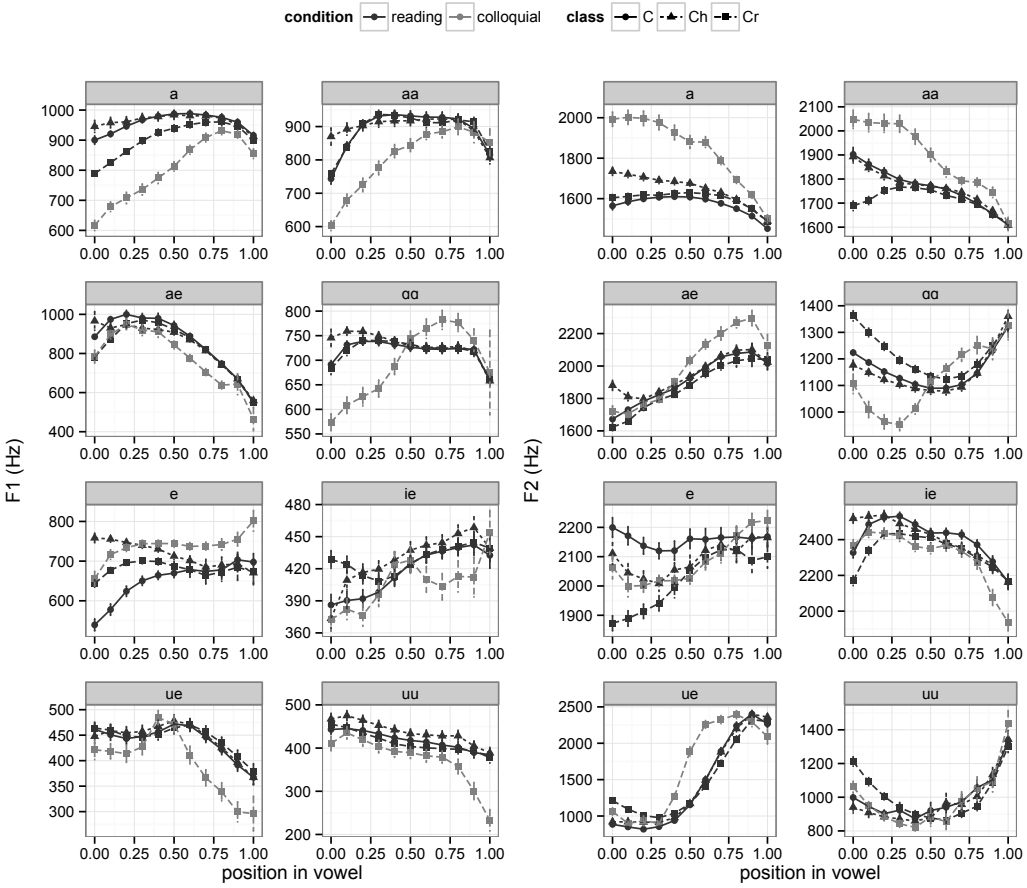
A second possibility is that the difference is again related to the prosodic environments in which the reading and colloquial items were collected. Several studies have found that duration of VOT can be modulated by accentual prominence (e.g. Cole et al., 2007; Cho & Keating, 2009). If onsets in colloquial /CrV/ clusters are in fact phonologically identical to obstruents in /C^{h}V/ contexts, the effects reported here would be consistent with a prosodically conditioned weakening account.

3.3.4. Vowel quality

Previous work on PP Khmer suggests that loss of /r/ conditions a process of low vowel raising in addition to f0 lowering. This finding was corroborated by the present study. As seen in Figure 5, F1 in items containing the low monophthongs /a aː ɑː/ is lower at nucleus onset in the `colloquial` condition than in the

`reading` condition. While F2 is higher at nucleus onset for /a aː/, consist with values for /e/ (around 2000 Hz), F2 is around 150 Hz lower at the onset of colloquial /ɑː/, consistent with values of /ɔ/. This also suggests that `reading` condition /ɑː/ may be realized as a more centralized [ɐː] (Wayland & Guion, 2005).

Figure 5: Average F1 (left) and F2 (right) trajectories in (/CV/, /C^{h}V/, /CrV/) `reading` vs. (/CrV/) `colloquial` condition by vowel nucleus type.



The magnitudes of these effects can be obtained by using syllable CLASS, vowel HEIGHT and BACKNESS and syllabic POSITION as covariates in a GLMM

predicting F1 (or F2). The vowels /a aː ɑː/ were coded as `[-high]`, and all other vowels as `[+high]`; similarly, /ɑː ɯe uː/ were coded as `[+back]` and other vowels as `[-back]`. To facilitate comparison between within-condition values, CLASS was coded as a factor with four levels: `/CV/`, `/ChV/`, `/CrV/` in `reading` condition, and `/CrC/` (representing /CrV/ syllables in the `colloquial` condition). Both models included three-way interactions between CLASS, POSITION, and HEIGHT and BACKNESS, along with random intercepts for subjects and items, correlated random by-subject slopes for POSITION and CLASS, and a random by-item slope for POSITION.

For models fit to F1, no significant interactions between syllable CLASS and POSITION for `[+high]` vowels were found, suggesting that the time course of F1 was similar across conditions for `[+high]` vowels in all syllable types. However, average F1 for `[-high]` vowels in /CrV/ contexts is estimated to be around 170 Hz lower in casual speech ($\beta = -172.21, SE = 9.06, t = -19, p < 0.001$). This effect was mediated by a three-way interaction between CLASS, POSITION and HEIGHT, indicating that F1 rises over the course of production for `[-high]` vowels in `colloquial` /CrV/ speech contexts ($\beta = 18.74, SE = 1.4, t = 13.7, p < 0.001$). A similar effect, though of lesser magnitude, was found for `[-back]` vowels in this same context ($\beta = 3.63, SE = 1.6, t = 2.3, p < 0.05$).

The coefficients for models fit to F2 data are slightly more challenging to interpret, as the effects do not pattern neatly by height or backness. Although there was a significant interaction between POSITION and HEIGHT for `[-high]` vowels in `colloquial` /CrV/ contexts ($\beta = -14.16, SE = 2.85, t = -5, p < 0.0001$), F2 for `[-high, -back]` /a aː/ falls over the course of the rime in `colloquial` /CrV/ syllables but actually rises for `[-high, +back]` /ɑː/ in

the same context (Figure 5). While F2 of [+back] vowels in general rises over the syllable rime ($\beta = 55.18, SE = 26.32, t = 2.1, p < 0.05$), F2 falls for both [−high, −back] /aː/ and [+high, −back] /ie/. Most important for present purposes is the fact that only terms where CLASS=/CrC/ (i.e., /CrV/ syllables in colloquial condition) emerged as significant. No evidence for monophthongization of diphthongs was observed in either condition.

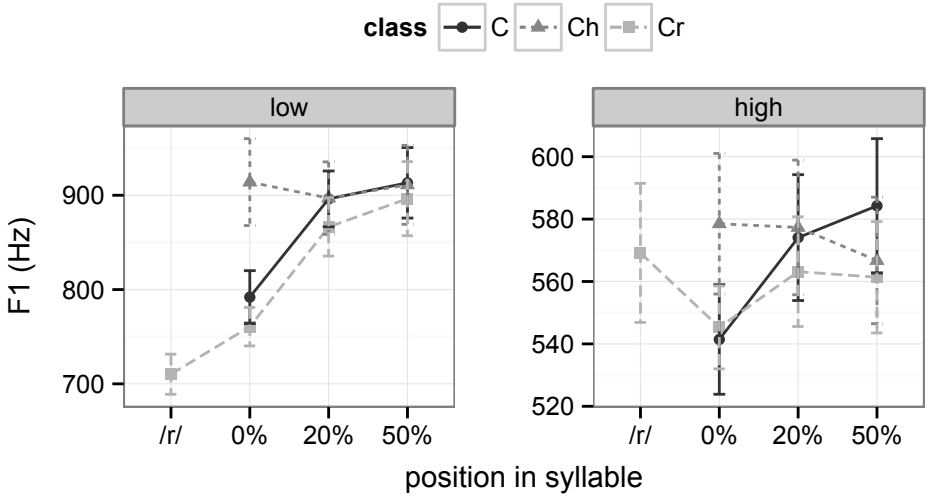
3.3.5. F1 drop

Wayland & Guion (2005) hypothesized that /r/ could exert a greater coarticulatory influence on tongue body height for low vowels than for mid or high vowels, thereby causing greater F1 lowering for low vowels in /CrV/ contexts compared to in /C^{h}V/ or /CV/ contexts. This hypothesis was assessed by examining the difference between F1 at voicing onset and points later in the rime (Figure 6). For low vowels, the F1 difference between voicing onset and a point 20% after nuclear vowel onset is significantly greater in /CrV/ contexts compared to /CV/ and /C^{h}V/ contexts ($\beta = -26.83, SE = 12.65, t = -2.1, p < 0.05$), but not for nonlow vowels ($p = 0.23$). However, this effect disappears if voicing onset for /CrV/ items is taken to be the onset of nuclear vowel ($p = 0.48$). A similar result obtains when measuring at 50% into the vowel; again, the difference is significant only for low vowels ($\beta = -31.40, SE = 13.65, t = -2.3, p < 0.05$), and only when onset of voicing for /CrV/ forms is taken to be the absolute onset of voicing rather than the onset of nuclear vowel voicing.

3.3.6. Voice quality (phonation type)

Items were also examined for evidence of a phonation type distinction by measuring the amplitude differential between the first and second harmonics (H1−H2)

Figure 6: Average F1 (in Hz) by vowel height and syllable type for `reading` condition items at four timepoints. Bars show standard error of the mean.



as well as the difference between the amplitude of the first harmonic and the amplitude of the most prominent harmonic of the third formant (H1−A3). Because raw amplitude differentials are known to be unreliable for data involving high f0 and/or high vowels, we instead report the measures H1*−H2* and H1*−A3*, corrected using the method described in Iseli & Alwan (2004).

Figure 7 compares the differences in these measures for /CV/, /CrV/, and /C^{h}V/ `reading` condition items with /CrV/ `colloquial` condition items. To get a more precise sense of the differences, we can examine the coefficients of GLMMs using backwards difference-coded CLASS (4 levels, with `/CV/` as reference level), POSITION, and their interaction (along with random intercepts for subjects and items) fit to H1*−H2* and H1*−A3*.

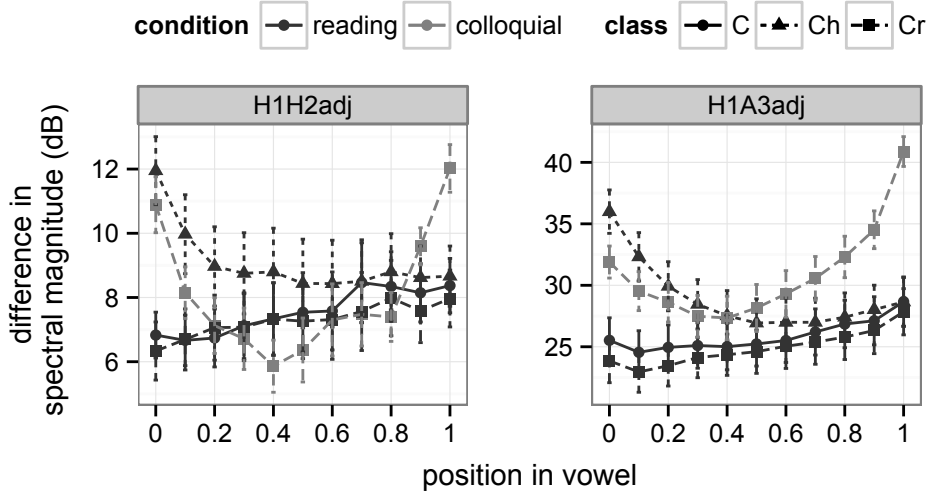
In neither model was the main effect of CLASS=`/CrV/` significant, indicating

no difference in the mean value of H1*−H2* or H1*−A3* between `reading` condition /CV/ and /CrV/ items. Both H1*−H2* ($\beta = 3.87, SE = 1.53, t = 2.5, p < 0.05$) and H1*−A3* ($\beta = 10.08, SE = 1.42, t = 7.1, p < 0.0001$) are greater on average in /C^{h}V/ than in /CrV/ (and, by extension, /CV/) `reading` condition contexts. For `colloquial` condition /CrV/ items, both measures are on average lower compared to /C^{h}V/ contexts (H1*−H2*: $\beta = -3.08, SE = 1.53, t = -2, p < 0.05$; H1*−A3*: $\beta = -6.16, SE = 1.42, t = 7.1, p < 0.0001$).

As can be seen in Figure 7, the value of both measures changes somewhat over the course of the syllable rime, especially in /C^{h}V/ and `colloquial` condition /CrV/ contexts. H1*−H2* in /C^{h}V/ items tends to fall ($\beta = -0.35, SE = 0.04, t = -9, p < 0.0001$) but rises overall in `colloquial` /CrV/ ($\beta = 0.34, SE = 0.04, t = 8.7, p < 0.0001$). A similar pattern is observed for H1*−A3*, i.e. falling in /C^{h}V/ contexts ($\beta = -0.99, SE = 0.05, t = -21.7, p < 0.0001$) but falling-rising in `colloquial` /CrV/ contexts ($\beta = 1.34, SE = 0.05, t = 29, p < 0.0001$). Interactions of POSITION and other syllable types were not significant.

The similarity in voice quality measures, especially at the syllable onset, between /C^{h}V/ and `colloquial` condition /CrV/ (> [C^{h}V̀]) forms raises the possibility that the acoustic effects observed may actually result from carryover aspiration, rather than a phonation type difference per se. To test whether spectral tilt values characteristic of breathy voice systematically co-occur with positive VOT, a GLMM was fit to predict reading condition H1*−A3* (averaged over repetitions) at the onset of voicing from the interaction of syllable CLASS and the DURATION of VOT (with random intercepts for subjects and items). While there is a main effect of CLASS=/`C`^{h}`V`/ at vowel onset ($\beta = 10.22, SE = 2.07, t = -4.9, p < 0.0001$), the effect disappears by vowel midpoint ($p = 0.11$). For H1*−H2*, nei-

Figure 7: H1*−H2* and H1*−A3* by CONDITION and CLASS, averaged across vowels, talkers, and repetitions. Bars indicate standard error of the mean.



ther the main effect nor the interaction are significant at vowel onset or offset, but similar effects are observed at vowel midpoint. Thus, while acoustic measures of breathy voice do bear some relationship to VOT, it does not seem to be the case that degree of breathy voicing is completely predictable from the temporal extent of aspiration: the differences in spectral balance induced by post-release aspiration in /C^{h}V/ forms do not persist throughout the vowel, whereas spectral balance differences (especially H1*−A3*) for `colloquial` /CrV/ forms do.

3.3.7. Realization of /r/

The realization of /r/ in the PP data varied considerably, both with respect to its presence/absence and its quality. Table 2 shows the rate of realization of /r/ by item in both conditions. The rate of 'accidental' realizations of /r/ in colloquial speech was greater for some items (e.g. /trap/ 'to imitate', /creh/ 'rust') than for others

(e.g. /krɑː/ ‘poor’). To get a sense of whether token frequency might be impacting the likelihood of /r/-dropping, frequency counts for these items were retrieved from the SBBIC Khmer frequency corpus (sbbic.org, 2011), which contains some 1.6 million tokens of over 115 000 Khmer wordforms. As seen in Table 2, the relationship between corpus frequency and /r/-realization is far from linear, but suggests that usage rates might play some role in conditioning realization of the trill.

The rate of /r/-loss also varies by speaker as well as by item. Table 3 shows the maximum likelihood estimates of /r/-realization based on the number of measurable /r/s produced by a given speaker in each condition. While some subjects are extremely consistent (e.g. s1, s6, s8), others are more variable, with some (e.g. s9) frequently omitting /r/s in reading condition forms and others (e.g. s12, s13, s19) frequently producing them in their colloquial pronunciations. Together with the variation observed in Table 2, it appears that at present, /r/-dropping (potentially accompanied by the use of lowered f0) in PP Khmer is highly speaker- and item-specific.

Table 2: Probability of /r/ realization in /CrV/ items by token frequency and condition, ranked by rate of colloquial realization. n gives the corpus frequency count from the SBBIC corpus.

	item	*read*	*colloq*	n		*item*	*read*	*colloq*	n
/krɑː/	‘poor’	0.91	0.19	383	/creh/	‘rust’	0.92	0.27	5
/priej/	‘spirit’	0.85	0.19	24	/craːp/	‘to shudder’	0.83	0.36	8
/crieŋ/	‘to sing’	0.89	0.23	284	/traː/	‘seal, stamp’	1.0	0.36	223
/kruː/	‘teacher’	0.85	0.23	435	/trap/	‘to imitate’	0.92	0.4	21
/prap/	‘to inform’	0.88	0.25	1947	/trae/	‘trumpet’	0.98	0.51	95

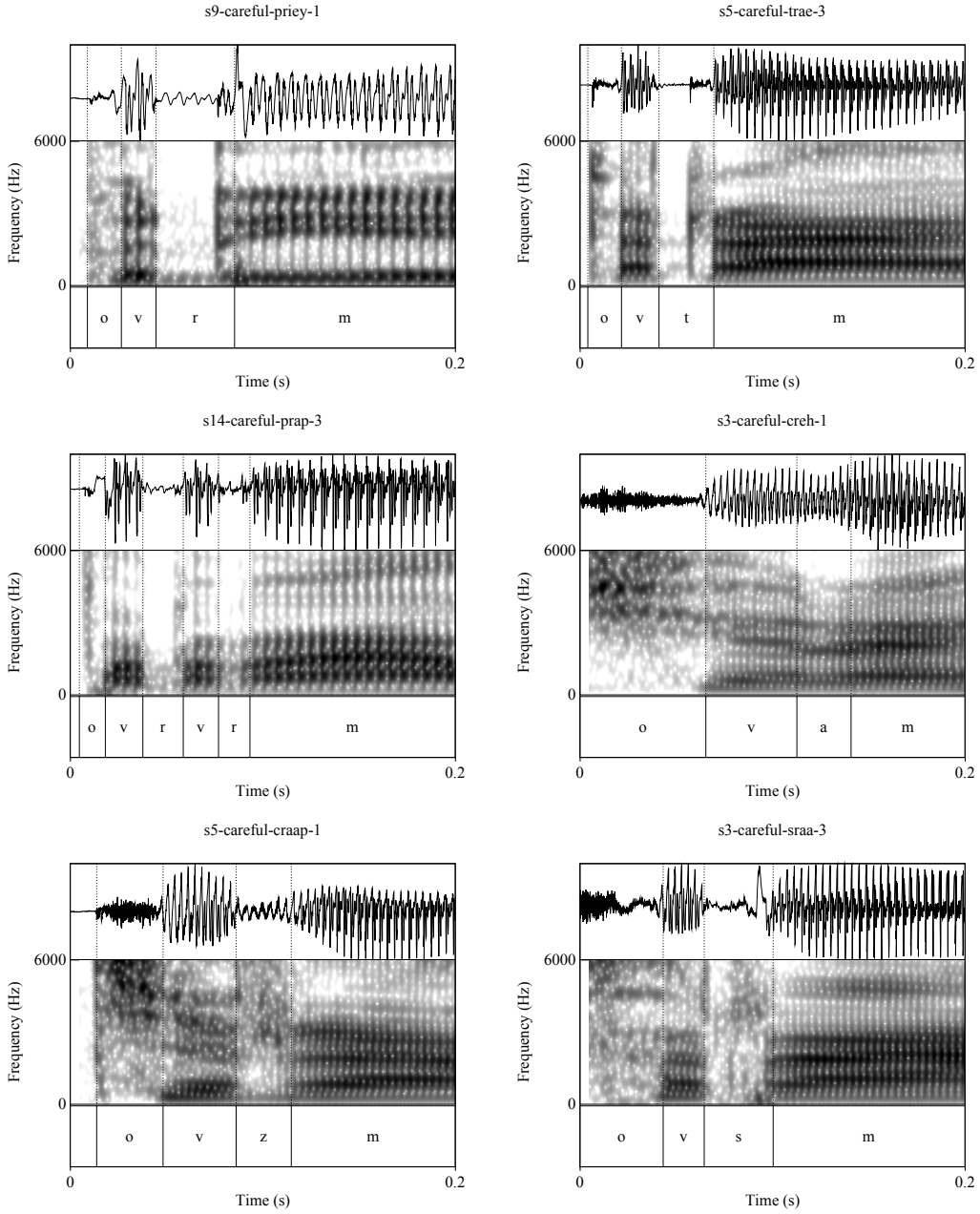
Table 3: Probability of trill realization in /CrV/ forms by speaker.

subj	*read*	*colloq*	*subj*	*read*	*colloq*
s1	1.0	0.3	s9	0.67	0.42
s2	1.0	0.18	s10	1.0	0.09
s3	0.97	0.24	s11	0.97	0
s4	1.0	0.27	s12	0.94	0.75
s5	0.45	0.24	s13	0.88	0.97
s6	1.0	0	s14	0.85	0.12
s7	1.0	0.06	s19	1.0	0.94
s8	1.0	0.09	s20	0.88	0.3

The phonetic realization of /r/ was also somewhat variable. Out of 527 `reading` condition /CrV/ tokens where some type of /r/ was realized, over 66% (350) were voiced during the closure, although the vast majority ($> 99\%$) of these were single taps, not trills. 53 instances (10%) of trills or taps were produced without measurable closure voicing. There were 59 instances (11%) of approximants, where there was no spectral evidence of complete closure. Finally, 65 (13%) tokens were realized as fricatives, 57 (12%) of which were voiced, with just 7 (1%) potentially voiceless fricatives.[6] Examples of each of these realizations is shown in Figure 8 (additional examples, together with audio recordings, are available as part of the online Supplementary Materials). These findings are consistent with the cross-linguistic tendency for trills to vary synchronically and diachronically with taps, approximants, and fricatives (Ladefoged & Maddieson, 1996; Solé, 2002).

[6]When present in `colloquial` condition forms, /r/ was always a voiced tap or trill.

Figure 8: Examples of variation in /ɹ/ realization, `reading` condition items. Row 1: voiced (left) and voiceless (right) taps. Row 2: voiced trill with two periods (left); approximant /ɹ/ (right). Row 3: voiced (left) and voiceless (right) fricatives. Labels are codes indicating segment type, not IPA transcriptions.



The tendency to produce noncanonical /r/ was more pronounced for some speakers than others (Table 4). For examples, speaker 10 produced over twice as many voiceless /r/s (19) as voiced variants (8), while speaker 5 produced nearly as many voiced fricatives (16) as other types of /r/ combined (17). Speakers 9 and 11 produced approximately one-third of their /r/s as approximants. For most participants, however, the voiced tap realization was dominant.

Table 4: Counts of /r/ realization in `reading` condition by speaker.

	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s19	s20
voiced /r/	28	29	20	26	10	22	22	14	19	8	20	31	25	32	19	25
voiceless /r/	4	1	3	6	2	1	3	8	2	19	0	0	2	1	1	0
approximant	0	0	3	1	5	5	6	6	11	1	11	1	5	0	2	2
voiced fric.	1	1	7	0	16	5	2	4	1	2	2	1	1	0	9	5
voiceless fric.	0	2	0	0	0	0	0	1	0	3	0	0	0	0	1	1

4. Perceptual study

In addition to collecting acoustic data, a pair of perceptual experiments were conducted in order to explore the degree to which PP Khmer speakers are able to use differences in acoustic realization to distinguish colloquial forms like /kruː/ > [kʰùː] 'teacher' from /kuː/ > [kuː] 'pair'. Experiment 1 focused on the role of f0, while also considering the effects of aspiration and breathiness. Experiment 2 focused on the role of F1, while also considering how F1 sensitivity might be modulated by f0, aspiration, and breathiness.

All participants who completed the production tasks also completed the perception tasks. The order of the two perception tasks was counterbalanced across subjects, but the perception tasks followed production tasks for all participants.

4.1. Experiment 1: f0 continuum

4.1.1. Materials and methods

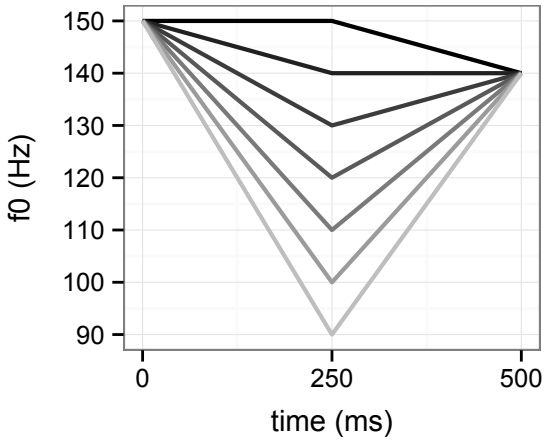
Experiment 1 focused on the perceptual salience of f0 as a cue to lexical identity of colloquial /CrV/ forms. By using a carrier syllable containing a high back vowel /u/, which has not been reported to differ in quality between reading and colloquial variants, the potentially confounding effects of F1 could be avoided (this dimension was addressed in Experiment 2).

4.1.2. Stimuli

The productions of a male speaker of Phnom Penh Khmer in his early 30s was used as a model for synthesis. He was born and raised in Phnom Penh and spent his entire life in Cambodia. The mean f0 contour of this speaker (who was not subsequently part of the experimental cohort) was computed based on 45 tokens (15 items $\times$ 3 repetitions); it was found to be broadly similar in shape (if not absolute value) to that of Speaker 1 reported in Wayland & Guion (2005): mean f0 at vowel onset was around 146 Hz, falling to around 105 Hz at vowel midpoint before rising again to around 135 Hz.

Using these productions and the data given in Wayland & Guion (2005) as an archetype, perceptual stimuli were generated using Praat's implementation of the KlattGrid synthesizer (Weenink, 2009). A 500 ms token of [kuː] was generated with f0 of 150 Hz at vowel onset and 140 Hz at vowel offset. Six additional tokens were generated by decreasing f0 at vowel midpoint (250 ms) in 10 Hz steps (0.95

Figure 9: Schematic f0 contours for synthesized stimuli in Experiment 1.



erb), with the lowest dip being to 90 Hz, yielding a seven step f0 continuum[7] (Figure 9).

In light of previous studies, VOT and breathy voice were also considered as potential cues. While the maximally thorough design would include similar continua for these two parameters, fully crossing three seven step continua would have resulted in nearly 350 unique stimuli before repetitions; given the participants' unfamiliarity with the type of task, this was judged to be too onerous, so a modified version of the procedure employed in Repp (1981) was used instead. The seven-step f0 continuum [kuː ∼ kùː] was used as the basis for two additional continua: one in which aspiration was increased uniformly by 70 ms in all seven items, yielding a continuum of [k^huː ∼ k^hùː], and one in which breathy voice was simulated by increasing breathiness amplitude (ATU) and spectral tilt (TL) to 75

[7]While 90 Hz was lower than the absolute lowest f0 value of the model speaker, it is consistent with the maximal drop of 60 Hz observed for speaker 1 reported in Wayland & Guion (2005).

and 24 dB, respectively, yielding a continuum of [ku̟ː ∼ kù̟ː]. This resulted in a total of 21 unique stimuli (7 f0 continuum steps × 3 conditions).

This design, where one acoustic parameter is varied quasi-continuously and the others categorically, allowed the potentially additive effect of a variety of acoustic cues to be explored with a minimum number of trials: by examining the shift in the identification function between, say, the [kʰuː ∼ kʰùː] and [kuː ∼ kùː] continua, a general idea of the influence (if any) of the categorical variable may be ascertained.

4.1.3. Participants

All participants in the perception study also completed the production tasks described above.

4.1.4. Procedure

Stimulus presentation and data collection were performed using Praat 5.2.26 (Boersma & Weenink, 2011); auditory presentation of stimuli was via Sony MDR-V600 headphones at a comfortable listening level. In each trial, participants responded via keyboard to indicate which of two lexical items (/kuː/ 'pair' or /kruː/ 'teacher') the stimulus most resembled. At no point were items represented orthographically (/r/ is a highly salient element of Khmer orthography); instead, pictures corresponding to the relevant lexical items were displayed on a laptop screen. A training set consisting of 10 repetitions of the two category exemplars ([kuː] and [kʰù̟ː]) was presented to each participant. Although a few participants required multiple passes through the training set, all were eventually able to correctly identify at least 80% of the category exemplars before continuing.

The 21 stimuli were presented in 10 randomized blocks for a total of 210 trials.

Reaction times and accuracies were recorded, and participants were allowed to take as much time as they wished to make a decision. There were no pauses during the experiment, which took approximately 10-15 minutes to complete, depending on the participants' rate of response.

4.1.5. Results

To gain a better understanding of the influences of stimulus step and condition on response choice, participant responses were modeled using GLMMs with a logistic link function. In general, most participants displayed a tendency for items with no medial f0 dip to be identified as /kuː/, with the rate of /kruː/ responses increasing with the extent of the f0 perturbation. This tendency is broadly true across conditions as well. However, two participants (s2 and s3) showed little or no sensitivity to this manipulation, and one participant (s19) appeared to respond at random.[8] Responses from these three participants were withheld from further analysis. In addition, participant responses with reaction times of less than 200ms as measured from onset of stimulus presentation (35 of 3570 responses, or 0.01%) were removed prior to model fitting and comparison; however, the main findings as reported below do not depend on either of these removals (though some coefficient estimates change slightly.)

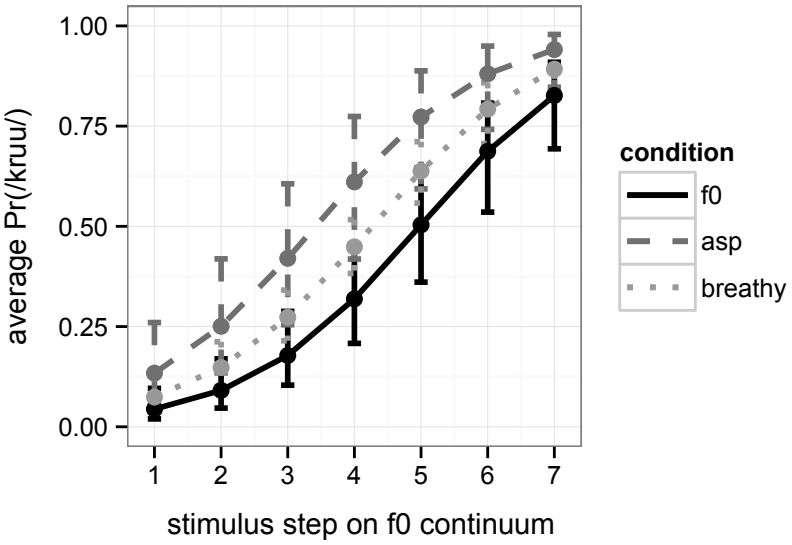
In the course of model comparison a variety of predictor variables were considered, including F0 STEP, CONDITION (with levels `f0`, `asp(irated)` and `breathy`, treatment coded with `f0` as the reference level), TRIAL, AGE, GENDER, and EDUCATION, along with various interactions between these predictors and by-participant random slopes and intercepts. However, likelihood-ratio tests

[8] See the online Supplementary Materials for details of the individual response patterns.

Table 5: Summary of fixed effects for GLMM fit to Experiment 1 data: coefficient estimates β, standard errors SE(β), Wald z-score ($= \beta/\text{SE}(\beta)$) and significance level p for all model predictors.

	Coef β	SE(β)	z	Pr($<\|z\|$)
(Intercept)	-0.76	0.29	-2.6	$<.01$
f0 step	0.77	0.09	8.6	$<.0001$
condition:asp	1.21	0.28	4.3	$<.0001$
condition:breathy	0.55	0.22	2.5	$<.05$

Figure 10: Average probability of /kruː/ responses in Experiment 1. Error bars indicate 95% average confidence intervals.



suggested that the more complex model structures were unwarranted: the final model included just two predictors (F0 STEP and CONDITION) along with by-participant random slopes and intercepts, which provided a significant improvement in log-likelihood over a model with random intercepts only ($\chi^2 = 233.33, p < 0.001$). When added, predictors such as TRIAL, AGE, GENDER, and EDUCATION did not reach significance, and their addition did not significantly improve model fit, nor did interaction terms or trial-specific random slopes.

Model coefficients and standard errors of fixed predictors are given in Table 5, with the predictions visualized (back-transformed into probability space) in Figure 10. A change in F0 STEP corresponds to a positive difference of 0.77 in the log probability of a /kruː/ response, or roughly speaking, a single change in F0 STEP corresponds maximally to a 20% shift in the probability of a /kruː/ response. Whether or not a token was aspirated also had a significant impact on its likelihood of being labeled /kruː/, as did breathiness, albeit to a lesser degree. Examination of the random effect estimates (not shown here) suggests that this result should be interpreted with caution, however, as the standard deviations of the random effect estimates for CONDITION are nearly the size of the fixed effect estimates, suggesting that participants varied considerably in their response to CONDITION. The small standard error of the random effect estimate of F0 STEP indicates greater uniformity in participant response to changes in this predictor.

4.2. Experiment 2: F1/F2 continuum

4.2.1. Materials and methods

Experiment 2 focused on the perceptual salience of vowel quality (primarily height) as a cue to lexical identity of colloquial /CrV/ forms. This experiment employed a stimulus containing a low back vowel /ɑː/, which has been reported

to diphthongize in colloquial production to [ɔɑ], in order to explore the perceptual salience of vowel quality as a cue to lexical identity.

4.2.2. Stimuli

The range of values for the stimuli in Experiment 2 were based on the same sources as for Experiment 1. A 500 ms token of the word [kɑː] 'neck' was again synthesized using the KlattGrid speech synthesizer, but here a seven step [kɑː ∼ kɔɑ] continuum was created by varying F1 and F2 at vowel onset in 20 and 25 Hz steps, respectively (i.e. the [kɑː] endpoint had F1 = 650 Hz, F2 = 1050 Hz, while the [kɔɑ] endpoint had F1 = 530 Hz, F2 = 900 Hz: see Figure 11). This manipulation is consistent with the realizations of low back /ɑː/ observed in section 3.3.4, where the mean difference in F1 between reading and colloquial /ɑː/ at vowel onset was found to be just over 90 Hz. Thus, as in Experiment 1, the extent of the manipulation extended past the hypothesized crossover point.

Using this continuum as a basis, two additional seven-step continua were created by uniformly increasing aspiration by 70 ms ([kʰɑː ∼ kʰɔɑ]) or increasing breathiness amplitude (ATU) and spectral tilt (TL) to 75 and 24 dB ([k̤ɑ̤ː ∼ kɔ̤ɑ̤]), as in Experiment 1. Uniformly dropping f0 60 Hz at vowel midpoint created a third [kɑ̀ː ∼ kɔ̀ɑ] continuum. This procedure resulted in a total of 28 unique stimuli (7 F1/F2 continuum steps × 4 conditions).

4.2.3. Participants

All participants in Experiment 1 also participated in Experiment 2.

4.2.4. Procedure

Procedures for Experiment 2 were largely identical to Experiment 1, except that here choices were /kɑː/ 'neck' or /krɑː/ 'poor'. 10 repetitions of the two

Figure 11: Schematic F1, F2, and f0 contours for synthesized stimuli in Experiment 2.

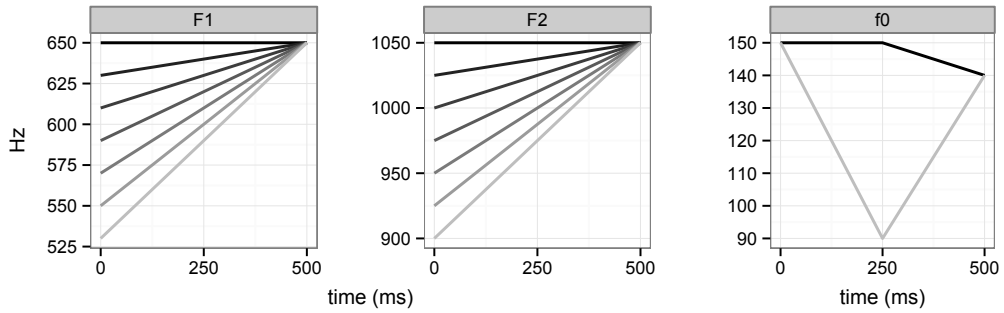
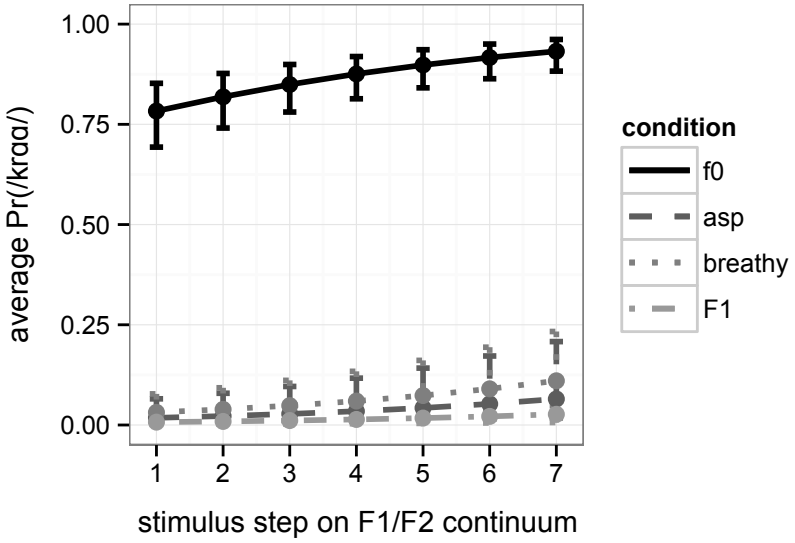


Table 6: Summary of fixed effects for GLMM fit to Experiment 2 data: coefficient estimates β, standard errors SE(β), Wald z-score ($= \beta/\text{SE}(\beta)$) and significance level p for all model predictors.

	Coef β	SE(β)	z	Pr($>|z|$)
(Intercept)	-4.26	0.73	-5.8	$<.0001$
F1/F2 step	0.22	0.04	5.3	$<.0001$
condition=f0	6.22	0.78	8.0	$<.0001$
condition=asp	0.93	0.39	2.4	$<.05$
condition=breathy	1.51	0.36	4.2	$<.0001$

category exemplars ([kɑː] and [k$^{\text{h}}$ɔ̤ɑ]) were presented to each participant to insure accurate identification of canonical stimulus endpoints. The 28 stimuli were then presented in 10 randomized blocks for a total of 280 trials. Reaction times and accuracies were recorded, though again participants were allowed to take as much time as they wished to make a decision.

Figure 12: Average probability of /krɑː/ responses in Experiment 2. Error bars indicate 95% average confidence intervals.



4.2.5. Results

Four participants (s12, s13, s18, and s19) appeared to be responding at random regardless of condition; responses from these participants were withheld from further analysis. Of the remaining participant responses, those with reaction times of less than 200ms as measured from onset of stimulus presentation (42 of 4480 responses, or 0.01%) were removed prior to model fitting and comparison; however, the main findings as reported below do not depend on either of these removals (though some coefficient estimates change slightly).

Participant responses for Experiment 2 were also modeled using multilevel logistic regression. The predictors included in the final model were F1/F2 STEP and CONDITION (treatment coded, with reference level `F1` plus `f0`, `aspirated`, and `breathy`). The final model included by-participant slopes and intercepts for

both predictors. As in Experiment 1, predictors such as TRIAL, AGE, GENDER, and EDUCATION did not reach significance and their addition did not significantly improve model fit, nor did the inclusion of interaction terms or trial-specific random slopes.

Model coefficients and standard errors of fixed predictors are given in Table 6, and the predictions visualized in Figure 12. While the estimates for the fixed predictors all reach statistical significance at the 0.05 level, the size of the estimated coefficients is much more telling: tokens with a medial f0 dip were almost certain to be identified as /krɑː/, while tokens from other conditions were generally labeled /kɑː/.

5. Discussion

5.1. Summary

Acoustic analysis of the production data showed clear differences between the phonetic realizations of /CrV/ forms in read and casual PP speech. Forms in casual speech contexts were realized with lower f0, increased VOT, and greater differences in spectral magnitude relative to forms in read speech. Low vowels in casual speech forms also were found to diphthongize. When present, the acoustic realization of /r/ was found to be rather variable, but most realizations were voiced taps: only around 13% of tokens were realized as fricatives, and just 10% of all tokens were realized without clear closure voicing.

The results of the perception studies confirm that f0 is indeed a salient cue to the lexical identity of /CrV/ forms in PP Khmer. While both aspiration and breathy voice appear to function as redundant perceptual cues, changes in F1/F2 (vowel height) did not appear to greatly effect response probabilities.

5.2. General discussion

While the acoustic analysis largely corroborates the observations of previous studies, especially those of Wayland & Guion (2005), there are also important differences. Perhaps most noticeably, the shape of the f0 contour in the present study was low or low-falling for many speakers. While it is possible that differences in the elicitation context may have resulted in read and casual speech forms being produced in different prosodic contexts, individual f0 patterns ranged from low level to low falling to falling-rising. Taken together, this suggests that the relevant aspect of pitch active in distinguishing these forms may simply be low versus non-low, rather than necessarily involving a contour.

A second difference observed in the present study was that the extent of post-release aspiration in colloquial /CrV/ forms was not as great as that in reading condition /C^{h}V/ forms, suggesting the possible existence of a 'semi-aspirated' phonetic category. This difference too could potentially be explained as an effect of prosodic structure, although further research would be necessary to confirm or deny this hypothesis.

Section 3.3.6 showed some evidence for colloquial /CrV/ items to be distinguished from reading condition items by measures of spectral tilt and open quotient. The effects were stronger for H1$^{*}-$A3* than for H1$^{*}-$H2*, but reading condition /C^{h}V/ and colloquial condition /CrV/ ($>$[C^{h}V̀]) forms were similar on both of these measures at many points in the vowel, particularly at the onset. While this suggests that the observed acoustic effects may actually be the result of carryover aspiration, rather than a phonation type difference, the degree of breathy voicing was not found to be redundantly predictable from the temporal extent of aspiration: differences in spectral balance induced by post-release aspiration in

/C^{h}V/ forms were not persistent throughout the vowel, whereas spectral balance differences for colloquial /CrV/ forms were.

Wayland & Guion (2005) proposed that the vowel quality change in low vowels may be a coarticulatory effect of the dorsal gesture associated with /r/, predicting greater F1 lowering in /CrV/ contexts. The results of the present study were consistent with this hypothesis, although the magnitude of the effect predicted by the model (section 3.3.4, Figure 6) is somewhat less than the F1 differences observed in the production data. Given the overall robustness of the diphthongization of low vowels, then, it is perhaps surprising that listeners did not respond to the F1/F2 manipulation in Experiment 2. This is unlikely to be an order effect, as half of participants completed Experiment 2 prior to Experiment 1. The degree of F1/F2 manipulation employed in Experiment 2 was greater than the comparable shift observed in production (section 4.2.2); if listeners do employ this cue, it might have been expected to at least have conditioned responses at one end of the continuum, but there was surprisingly little between-participant variability on this point. The fact that [ɔɑ], the diphthong that surfaces in colloquial PP pronunciations of /kraː/, is not part of the otherwise substantial inventory of Khmer diphthongs would presumably *increase* its availability to function as a cue to lexical identity. Similar tests with a wider range of diphthongs are necessary to resolve this issue.

5.3. Tonogenetic mechanisms

To explain the emergence of f0 as a cue to lexical identity in colloquial PP /CrV/ forms, Wayland & Guion (2005) proposed a new tonogenetic mechanism: phonologization of a drop in f0 conditioned by a combination of (a) the high degree of airflow necessary to maintain trilling and (b) subsequent devoicing of

the trill. The perceptual results (section 4.1), demonstrating the salience of the f0 drop in found in /CrV/ > [C^{h}V̀] forms, are consistent with this hypothesis. In production, however, the difference in f0 between /r/ and points in the nucleus of /CrV/ items was not significantly different from the difference between f0 at voicing onset and points in the nucleus of /CV/ or /C^{h}V/ items (section 3.3.2). Although this finding is not strictly incompatible with the acoustic-aerodynamic hypothesis, it raises the question of why a perceptually salient phonetic precursor would be phonologized in one context but not in others where it is of similar magnitude (cf. Kirby, 2013).

The alternative hypothesis suggested in the Introduction is that the common phonetic precursor behind the f0 lowering, F1 raising and aspiration observed in PP Khmer was a stage of breathy phonation. This is based in the long-standing idea that tone and register systems evolve through a stage of contrastive voice quality. In this regard, it is instructive to consider Huffman's (1985) model of the diachronic paths that have led to the development of register and tone systems in the Mon-Khmer family (Table 7; cf. Ferlus, 1980). Huffman posits a proto-contrast between voiced and voiceless initials, evidence for which can be found both in 'conservative' Mon-Khmer languages such as Eastern Kammu (Premsrirat, 2001) as well as in the Khmer writing system (Pinnow, 1957). In the 'transitional' stage, voiced stops develop into breathy-voiced stops and then voiceless aspirated stops. Systems in the 'register' stage are those which contrast breathy-voiced with modally-voiced vowels; the breathy voiced register may in turn condition F1 lowering, creating high onglides to the low vowel series, as took place in Middle Khmer between the 16th and 18th centuries (Jenner, 1974). Alternatively, instead of 'restructuring', the phonation type distinctions may transphonol-

ogize into an f0 contrast, with vowels in the breathy register bearing low or falling tone, as evidenced by e.g. Western Kammu dialects (Svantesson & House, 2006; Abramson et al., 2007).[9]

Table 7: Diachronic paths leading to register/tone systems. After Huffman (1985).

	Proto-language	Conservative	Transitional	Register	Restructured	(Tonal)
2nd	*gaa	gaa	k^{h}aa	kạ̀a	kia	(kàa)
1st	*kaa	kaa	kaa	kaa	kaa	(káa)

The phonetic motivations for this proposal are grounded in the observation that breathy phonation frequently co-occurs with high vowels (Henderson, 1952; Huffman, 1976; Denning, 1989). This correlation has a plausible acoustic-perceptual basis, in that the perceptual integration of low-frequency components in breathy vowels with F1 conditions the percept of a 'higher' vowel (Lotto et al., 1997). The

[9] A reviewer raises the case of Chanthaburi Khmer, which appears to have stopped at the intermediate stage between Huffman's 'register' and 'restructured' systems. In this dialect, there is (some) acoustic evidence for a contrast between modal and breathy voicing, with minimal pairs such as /kat/ 'to cut' vs. /kạt/ 'he' (Wayland & Jongman, 2003). If breathiness is meant to condition F1 raising, it is perhaps curious that this vowel has resisted diphthongization. However, the mere *presence* of a phonetic precursor does not entail that that precursor will inevitably be phonologized. Furthermore, it is not clear from Wayland & Jongman (2003) if this phenomenon is restricted to short /a/ only, to all low vowels, or to the particular set of lexical items examined therein. It seems reasonable to think that the diphthongization seen in PP Khmer may have required the combination of both coarticulatory and acoustic precursors, e.g. if the magnitudes of the individual precursors were insufficient to condition raising on their own.

articulatory configuration of a raised tongue body and a lowered larynx is common to both high vowels and breathy voicing, in addition to being associated with lowered fundamental frequency (Laver, 1980).[10] Thus, one could imagine that, over time, listeners-turned-speakers may come to produce breathy voiced forms with lower F1 and f0. The key difference between the PP Khmer case and Huffman's idealized pathway from voicing contrast to tone or register contrast would lie in the catalyst: in this case, variation in the realization of /r/, which may have initiated something like the 'transitional' stage without first passing through the 'conservative' stage. In all other respects, however, the data presented here are consistent with the hypothesis of voice quality-driven restructuring.

Both of the tonogenetic proposals discussed here rely to a greater or lesser extent on fortition of /r/. Wayland & Guion's proposal crucially appeals to trill devoicing, for which the present study found some limited evidence. The breathiness account does not strictly rely on devoicing, but on the presence of frication of any kind that might plausibly induce a breathy percept on the following vowel (Klatt & Klatt, 1990). In terms of /r/-variability, devoicing of /r/ was fairly rare, while fricativization was somewhat more common. Taken together with the differences in voice quality, this could be seen as slightly favoring a perceptually-driven explanation.

[10] At the same time, the intrinsic f0 of high vowels is known to be universally higher than that of low vowels (Whalen & Levitt, 1995), a fact thought to be related to the mechanics of tongue raising (Whalen et al., 1999). While clearly perceptible, however, vowel-intrinsic f0 differences are typically perceived as differences in vowel quality, not changes in pitch (Silverman, 1987; Fowler & Brown, 1997), and there are only a few putative cases of such differences conditioning the emergence of (or even interacting with) lexical tone (Hombert, 1977; Kingston, 2011).

If the evidence reviewed here does not seem to provide unassailable support for one position or the other, it should be kept in mind that arguments for the phonetic precursor(s) of a sound change based on the analysis of synchronic data will always be at best indirect. Thus, the fact that breathiness and/or fricativization of /r/ in reading condition PP forms is found only sporadically does not mean that these processes were not at one time more widespread; indeed, this study found considerable variation in the synchronic realization of /r/ in standard speech (section 3.3.7). While the PP Khmer speakers who participated in this study controlled both standard and colloquial registers, this sound change was presumably actuated by members of the speech community who spoke only the demotic. Moreover, the kind of perceptual hypo-correction leading to sound change may be more likely when the relevant variation is relatively rare, as listeners are less likely to have the experience necessary to compensate for predictable variants (Ohala, 1993).

Finally, it is worth noting that while the development of a distinctive pitch contour via /r/-loss seems to be unique to Khmer, it does not be appear to be localized to Phnom Penh. Thạch Ngọc Minh (1999) reports a shift of /r/ $>$ /h/ in the variety of Khmer spoken in and around Kiên Giang province in Vietnam.[11] Similar to Noss (1966), this shift is reported in absolute initial position only; however, the resulting pitch contour is reported to be falling, as found here, rather than falling-rising. Loss of /r/ in /CrV/ contexts similarly triggers a falling pitch, but

[11] A similar phenomenon is observed with /r/-initial forms in PP Khmer, e.g. /rien/ 'to learn'. Although not analyzed here, these forms also show evidence of f0 difference and a shift from /r/ $>$ [h], e.g. [hìen]. Assuming the acoustic characteristics of /rV/ syllables are similar to those of /CrV/ syllables, the same type of explanation, together with a phonotactic requirement for all Khmer syllables to have onsets, could also explain this shift in absolute initial position.

unlike in the PP case, no increase in aspiration is transcribed. Data on this pattern in Kiên Giang Khmer have also been collected and are currently being prepared for publication.

6. Conclusions

This study has established that loss of /r/ is associated with several sound changes in colloquial PP Khmer, particularly the development of increased aspiration, a low-falling or falling-rising f0 contour, and low vowel diphthongization. As demonstrated in sections 4.1 and 4.2, the f0 cue in particular is highly salient to native listeners, who are able to employ it to distinguish between lexical items. Here, I have suggested that a new tonogenetic mechanism is not necessary to explain this change; instead, the emergence of the f0 contrast may have been driven primarily via breathiness conditioned by variation in the realization of /r/, while low vowel diphthongization arose through a combination of acoustic and articulatory factors. This allows us to understand this seemingly unique sound change in terms of more widespread areal and cross-linguistic tendencies.

Acknowledgements

This research was funded in part by a Council of American Overseas Research Centers (CAORC) Senior Research Fellowship from the Center for Khmer Studies. Special thanks to Mr. Sor Sokny and the Buddhist Institute of Phnom Penh, and to all of the participants, without whom this work would not have been possible. Portions of this paper were presented at the University of Cologne, the 2nd Workshop on Sound Change, the MPI Leipzig Workshop on Tone: Theory and Practice, and the 22nd Meeting of the Southeast Asian Linguistics Society; I am

grateful to those audiences for their insightful and thought-provoking comments. Thanks also to Art Abramson, Marc Brunelle, Doug Cooper, Justin Watkins, Alan Yu, three anonymous reviewers, and Prof. Taehong Cho for additional advice, assistance, and commentary.

This paper is dedicated to the memory of Susan Guion Anderson, whose work with Ratree Wayland on Khmer tonogenesis provided the inspiration for this project.

Supplementary materials

The Praat scripts used to generate experimental stimuli used in this study, along with additional plots, audio examples, measurement files and accompanying R code, are available as part of the Supplementary Materials on the author's website at `http://www.lel.ed.ac.uk/~jkirby/khmer`.

Appendix A: Wordlist

pap	‘crack (of whip)’	p^{h}ap	‘bolt of cloth’	prap	‘to inform’
piɐj	‘extinct’	p^{h}iɐj	‘to spread out’	priɐj	‘spirit’
puəj	‘to swoop’	p^{h}uəj	‘blanket’	pruəj	‘sad’
tap	‘hurridly’	t^{h}ap	‘wallpaper’	trap	‘to imitate’
taɐ	‘only, just’	t^{h}aɐ	‘to take care of’	traɐ	‘trumpet’
taː	‘grandfather’	t^{h}aː	‘to say, tell’	traː	‘seal, stamp’
caːp	‘sparrow’	c^{h}aːp	‘to set on fire’	craːp	‘to shudder’
ceh	‘to know’	c^{h}eh	‘burning’	creh	‘rust’
cieŋ	‘more than’	c^{h}ieŋ	‘learning’	crieŋ	‘to sing’
kuː	‘pair’	k^{h}uː	‘old’	kruː	‘teacher’
kɑː	‘neck’	k^{h}ɑː	‘kind of soup’	krɑː	‘poor’
hien	‘to dare’			rien	‘to learn’
hɨŋ	‘altar’			rɨŋ	‘hard, solid’
sok	‘health’			srok	‘district’
saː	‘to gather’			sraː	‘alcohol’

Appendix B: GLMMs

In classical linear regression, a response variable y is modelled as a linear combination of k predictors plus a normally distributed error term ϵ, as in (1):

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon_i, \\ \epsilon_i &\sim N(\mu_i, \sigma^2) \end{aligned} \tag{1}$$

A GENERALIZED LINEAR MODEL (GLM) extends this framework to cover a broad range of commonly encountered types of dependent variables and error structures. A GLM consists of a linear predictor η along with a link function ℓ describing how the mean $E[y] = \mu$ depends on η, and a variance function V describing how the variance $Var[y]$ depends on $E[y]$:

$$\begin{aligned} \eta &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \\ \ell(\mu) &= \eta, \\ Var[y] &= V(\mu)\phi \end{aligned} \tag{2}$$

where ϕ is a (possibly unknown) scale parameter. It is convenient if the distribution of y_i is a member of the exponential family, but this is not strictly necessary. Classical linear regression with normally distributed error (1) is thus a special case of (2) where the link function is the identity function, and the variance function $V(\mu) = 1$. For logistic regression, where response data are binomial, $\ell = \text{logit}(\mu)$ and $V = \mu(1 - \mu)$.

In a GENERALIZED LINEAR MIXED MODEL (GLMM, also called a multilevel or 'mixed-effects' model), the effects of predictors are allowed to vary across clusters by including 'random' (i.e., cluster-specific) slopes and/or intercepts. For

example, in a model that includes random intercepts u_j for each participant j,

$$\begin{aligned} y_{ij} = \ & \beta_0 + u_{0j} + \beta_1 x_{ij} + \cdots + \beta_k x_{ij}, \\ & u_j \sim N(0, \sigma_u^2) \end{aligned} \tag{3}$$

the overall intercept for a given participant depends on the estimate of u_j, but the effect of the predictors $\beta_1, \ldots, \beta_k$ are assumed to be the same across participants. It is also possible to formulate a model that allows the effect of predictors to vary across participants, i.e.:

$$\begin{aligned} y_{ij} = \ & \beta_0 + \beta_1 x_{1ij} + \cdots + \beta_k x_{kij} + \\ & u_{0ij} + u_{1j} x_{1ij} + \cdots + u_{kj} x_{kij}, \\ & \mathbf{u}^{\mathrm{T}} = (u_{0j}, \ldots, u_{kj}) \sim MVN(\mathbf{0}, \mathbf{\Omega}_u) \end{aligned} \tag{4}$$

where $\mathbf{\Omega}_u$ is the covariance matrix of the multivariate normally-distributed (MVN) random effects. In this type of model, both the magnitude (intercept) and the direction (slope) of the effect of a predictor are allowed to vary for each participant. The model may then be extended to include multiple cluster types (e.g., items in addition to participants).

Agresti (2002) and Gelman & Hill (2007) provide statistical introductions to GLMMs. Some examples of use in language research include Baayen et al. (2008); Jaeger (2008); and Kong et al. (2011).

References

Abramson, A. S., Nye, P. W., & Luangthongkum, T. (2007). Voice register in Khmu': experiments in production and perception. *Phonetica*, *64*, 80–104.

Agresti, A. (2002). *Categorical data analysis*. (2nd ed.). New York: John Wiley & Sons.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. http://CRAN.R-project.org/package=lme4. R package version 0.999999-2.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–100.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.26, retrieved 17 June 2011 from http://www.praat.org/.

Cho, T. (2011). Laboratory phonology. In N. C. Kula, B. Botma, & K. Nasukawa (Eds.), *The Continuum Companion to Phonology* (pp. 343–368). London: Continuum.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.

Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: evidence from Radio News speech. *Journal of Phonetics*, *35*, 180–209.

Denning, K. (1989). *The diachronic development of phonological voice*. Ph.D. dissertation, Stanford University.

Diffloth, G. (1989). Proto-Austroasiatic creaky voice. *Mon-Khmer Studies*, *15*, 139–154.

Diffloth, G. (2003). Khmer. In W. Frawley (Ed.), *International encylopedia of linguistics, vol. II* (pp. 355–359). (2nd ed.).

Ferlus, M. (1980). Formation des registres et mutations consonantiques dans les langues Mon-Khmer. *Mon-Khmer Studies*, *8*, 1–76.

Ferlus, M. (1992). Essai de phonétique historique du khmer. *Mon-Khmer Studies*, *21*, 57–89.

Filippi, J.-M., & Vicheth, H. C. (2009). *Dictionnaire de la pronunciation du Khmer: Khmer standard et dialecte phnompenhois*. Phnom Penh: Editions Funan.

Fowler, C. A. & Brown, J. M. (1997). Intrinsic F0 differences in spoken and sung vowels and their perception by listeners. *Perception and Psychophysics*, *59*, 729–738.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gick, B. A., Ming Kang, & Whalen, D. H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics*, *30*, 357–371.

Gick, B. A., Campbell, F., Oh, S., & Tamburri-Watt, L. (2006). Towards universals in the gestural organization of syllables: a cross-linguistic study of liquids. *Journal of Phonetics*, *34*, 49–72.

Guion, S. G., & Wayland, R. P. (2004). Aerodynamic coarticulation in sound change or how onset trills can condition a falling tone. In A. Agwuele, W. Warren, & S.-H. Park (Eds.), *Proceedings of the 2003 Texas Linguistic Society Conference* (pp. 107–115). Sommerville, MA: Cascadilla Proceedings Project.

Headley, R. K., Chim, R., & Seoeum, O. (1997). *Modern Cambodian-English Dictionary*. Kensington, MA: Dunwoody Press.

Henderson, E. J. A. (1952). The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies*, *14*, 149–174.

Hombert, J.-M. (1975). *Towards a theory of tonogenesis: An empirical, physiologically and perceptually-based account of the development of tonal contrasts in language*. Ph.D. dissertation, University of California at Berkeley.

Hombert, J.-M. (1977). Development of tones from vowel height? *Journal of Phonetics*, *5*, 9–16.

Huffman, F. E. (1967). *An outline of Cambodian grammar*. Ph.D. dissertation, Cornell University.

Huffman, F. E. (1970). *Cambodian system of writing and beginning reader*. New Haven: Yale University Press.

Huffman, F. E. (1972). The boundary between the monosyllable and the disyllable in Cambodian. *Lingua*, *29*, 54–66.

Huffman, F. E. (1976). The register problem in fifteen Mon-Khmer languages. In P. N. Jenner, L. C. Thompson, & S. Starosta (Eds.), *Austroasiatic studies, part I* (pp. 575–589). Honolulu: The University Press of Hawaii.

Huffman, F. E. (1985). Vowel permutations in Austroasiatic languages. In *Linguistics of the Sino-Tibetan area: The state of the art: Papers presented to Paul K. Benedict for his 71st birthday* (pp. 141–145). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.

Iseli, M., & Alwan, A. (2004). An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation. In *Proc. ICASSP* (pp. 669–672). Montreal.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jenner, P. (1974). The development of registers in Standard Khmer. In N. D. Liem (Ed.), *South-east Asian Linguistic Studies Vol. 1* (pp. 47–60). Canberra: Pacific Linguistics.

Kingston, J. (2011). Tonogenesis. In M. van Oostendorp & C. J. Ewan & E. V. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 2304–2333). Oxford: Wiley-Blackwell.

Kirby, J. P. (2013). The role of probabilistic enhancement in phonologization. In A. Yu (Ed.), *Origins of sound patterns: approaches to phonologization* (pp. 228–246). Oxford: Oxford University Press.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.

Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, *39*, 196–211.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Cambridge: Blackwell.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on F0 in Taiwanese. In J. Brooke et al. (Eds.), *Proceedings of the 2nd International Conference on East Asian Linguistics*. Volume 2.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press: Cambridge.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 157–233). Cambridge, MA: MIT Press.

Lotto, A. J., Holt, L. L., & Kluender, K. R. (1997). Effect of voice quality on perceived height of English vowels. *Phonetica*, *54*, 76–93.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant types and tone* (pp. 71–95). Los Angeles: University of Southern California.

McGowan, R. S. (1992). Tongue-tip trills and vocal tract wall compliance. *Journal of the Acoustical Society of America*, *91*, 2903–2910.

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., & Kunnari, S. (2012).

Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics*, *40*, 796–807.

Noss, R. B. (1966). The treatment of */r/ in two modern Khmer dialects. In N. H. Zide (Ed.), *Studies in Comparative Austroasiatic Linguistics* (pp. 89–95). London: Mouton & Co.

Ohala, J. J. (1973). The physiology of tone. In L. Hyman (Ed.), *Consonant types and tone* (pp. 1–14). Los Angeles: University of Southern California.

Ohala, J. J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical Linguistics: Problems and Perspectives* (pp. 237–278). London: Longman.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.

Pinnow, H.-J. (1957). Sprachgeschichtliche Erwägungen zum Phonemesystem des Khmer. *Zeitschrift für Phonetik*, *10*, 378–391.

Pinnow, H.-J. (1980). Reflections on the history of the Khmer phonemic system. *Mon-Khmer Studies*, *7*, 103–130.

Pisitpanporn, N. (1994). On the r $>$ h shift in Phnom Penh Khmer. *Mon-Khmer Studies*, *24*, 105–113.

Pisitpanporn, N. (1999). A note on colloquial Phnom Penh Khmer. In G. Thurgood (Ed.), *Papers from the Ninth Annual Meeting of the Southeast Asian Linguistics Society* (pp. 243–248). Arizona State University: Program for Southeast Asian Studies.

Pittayaporn, P. (2007). Prosody of final particles in Thai: Interaction between lexical tones and boundary tones. Poster presented at the International Workshop on Intonational Phonology: Understudied or Fieldwork Languages. Retrieved 8 May 2013 from http://www.linguistics.ucla.edu/people/jun/Workshop2007ICPhS/Papers/Pittayaporn-Thai-Poster.pdf.

Premsrirat, S. (2001). Tonogenesis in Khmu dialects of SEA. *Mon-Khmer Studies*, *31*, 47–56.

Pulleyblank, E. G. (1978). The nature of Middle Chinese tones and their development to Early Mandarin. *Journal of Chinese Linguistics*, *6*, 173–203.

Repp, B. H. (1981). Phonetic and auditory trading relations between acoustic cues in speech perception: preliminary results. *Haskins Laboratories Status Report on Speech Research SR-67/68*, 165–189.

sbbic.org (2011). Khmer frequency corpus. Retrieved 5 Nov 2012 from http://sbbic.org/Khmer-Corpus-Work.zip.

Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Linguistics Laboratory*, *3*, 83–109.

Silverman, K. E. A. (1987). *The structure and processing of fundamental frequency contours*. Ph.D. dissertation, University of Cambridge.

Solé, M.-J. (2002). Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics*, *30*, 655–688.

Svantesson, J.-O., & House, D. (2006). Tone production, tone perception and Kammu tonogenesis. *Phonology*, *23*, 309–333.

Thạch Ngọc Minh (1999). Monosyllabization in Kiengiang Khmer. *Mon-Khmer Studies*, *29*, 81–95.

Thurgood, G. (2002). Vietnamese and tonogenesis: revising the model and the analysis. *Diachronica*, *19*, 333–363.

Wayland, R., & Jongman, A. (2002). Registrogenesis in Khmer: A phonetic account. *Mon-Khmer Studies*, *32*, 101–115.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics*, *31*, 181–201.

Wayland, R. P., & Guion, S. G. (2005). Sound changes following the loss of /r/ in Khmer: a new tonogenetic mechanism? *Mon-Khmer Studies*, *35*, 55–82.

Weenink, D. (2009). The KlattGrid speech synthesizer. In *Proc. Interspeech 2009* (pp. 2059–2062). Brighton, UK.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*, 349–356.

Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1999). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27*, 125–142.